BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 9/24, 9/26, 9/42

A1

(11) International Publication Number: WO 00/70031

(43) International Publication Date: 23 November 2000 (23.11.00)

(21) International Application Number: PCT/US00/13971

(22) International Filing Date: 19 May 2000 (19.05.00)

(30) Priority Data:
60/134,925 19 May 1999 (19.05.99) US

(71) Applicant *(for all designated States except US)*: MIDWEST RESEARCH INSTITUTE [US/US]; 425 Volker Boulevard, Kansas City, MO 64110 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HIMMEL, Michael, E. [US/US]; 9202 W. Hialeah Place, Littleton, CO 80123 (US). ADNEY, William, S. [US/US]; 13190 W. 21st Avenue, Golden, CO 80401 (US). BAKER, John, O. [US/US]; 18790 W. 60th Avenue, Golden, CO 80403 (US). VINZANT, Todd, B. [US/US]; 16601 W. 15th Avenue, Golden, CO 80401 (US). THOMAS, Steven, R. [US/US]; 485 Kearney Street, Denver, CO 80220 (US). SAKON, Joshua [US/US]; 1753 Zion Road, #82, Fayetteville, AR 72703 (US). DECKER, Stephen, R. [US/US]; 8020 Greenwood Drive, Berthod, CO 80403 (US).

(74) Agent: WHITE, Paul, J.; National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: E1 ENDOGLUCANASE VARIANTS Y245G, Y82R AND W42R

(57) Abstract

The invention provides a method for making a glycosyl hydrolase characterized by an increase in catalytic activity on an insoluble substrate, comprising replacing an active site associated glycosyl–stabilizing amino acid of the hydrolase with an amino acid, the replacing amino acid not strongly binding a disaccharide product in the active site, and a method for making a glycosyl hydrolase characterized by an increase in catalytic activity on a soluble substrate, comprising replacing a hydrophobic substrate binding amino acid of the hydrolase with a positively charged amino acid. The invention further provides a glycosyl hydrolase, comprising Y42R (SEQ. ID NO:1), W82R (SEQ. ID NO:2), or Y245G (1) (SEQ. ID NO:3) and the DNA sequences encoding the enzymes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

E1 ENDOGLUCANASE VARIANTS Y245G, Y82R AND W42R.

The United States Government has rights in this invention pursuant to Contract No. DE-AC36-99GO-10337 between the United States Department of Energy and the Midwest Research Institute.

Technical Field.

The present invention relates to glycosyl hydrolases, and in particular to variants of *Acidothermus cellulolyticus* E1 endoglucanase which demonstrate increases in catalytic activity on insoluble or soluble substrates.

Background Art.

Plant biomass, which represents the cellulosic materials that compose the cell walls of all higher plants, is the most abundant source of fermentable carbohydrates in the world. When biologically converted to fuels, such as ethanol, and various other low-value high-volume commodity products, this vast resource can provide environmental, economic and strategic benefits on a large scale, which are unparalleled by any other sustainable resource. See, Lynd L.R, *et al.*, *Science* 1991, 251: 1318-23; Lynd L.R, *et al.*, *Appl. Biochem. Biotechnol.* 1996, 57/58:741-61.

Cellulase enzymes provide a key means for achieving the tremendous benefits of biomass utilization, in the long term, because of the high sugar yields, which are possible, and the opportunity to apply the modern tools of biotechnology to reduce costs. However, the soluble products, cellobiose and glucose, have been reported to be powerful inhibitors of the cellulase complex and of the individual enzyme components: endoglucanase; cellobiohydrolase; and β-D-glucosidase. Howell J.A. *et al.*, *Biotechnol. Bioeng.*, 1975, XVII: 873.

The surface chemistry of acid pretreated-biomass, used in bioethanol production, is different from that found in native plant tissues, naturally digested by bacterial and fungal cellulase enzymes, in two important ways: (1) pretreatment heats the substrate past the phase-transition temperature of lignin; and (2) pretreated biomass contains less acetylated hemicellulose. Kong F., *et al.*,: *Appl. Biochem. Biotechnol.*, 1993, 34/35:23-35; *Handbook on Bioethanol: Production and Utilization.* edited by Wyman C.E., Washington, DC: Taylor & Francis, 1996: 424. Thus, it is believed that the cellulose fibers of pretreated-biomass, the objective of cellulase action, are embedded in a polymer matrix different from that of naturally occurring plant tissue. Therefore, for the efficient production of ethanol from pretreated

biomass, it is critical to improve the effectiveness of naturally occurring enzymes on that substrate, recognizing that nature may not have optimized mechanisms for enzymatic hydrolysis of such man-made substrates. A need therefor exists for modified cellulase enzymes. which are characterized by an increase in catalytic activity on either pure, or the cellulose component in a pretreated biomass.

Cellulases are modular enzymes composed of independently folded, structurally and functionally discrete domains. Typically, cellulase enzymes comprise a catalytic domain, comprised of active site residues, and one or more cellulose-binding domains, which are involved in anchoring the enzyme to cellulose surfaces. There are 21 families of catalytic domains, and each are classified on the basis of similarity of their amino acid sequences. The three-dimensional structure of 14 of those enzymes has been determined. These families exhibit a diverse range of folding patterns, but each maintains a conserved catalytic cleft. Cellulose hydrolysis is accompanied by either inversion or retention of the configuration of the anomeric carbon. Generally, for the retaining enzymes, the leaving group is the non-reducing side of cellulose. Whereas, for inverting enzymes, the leaving group is the reducing side of the cellulose. Although the folding pattern of the catalytic domains and the precise mechanisms of hydrolysis vary, their active site features remain similar. All catalytic clefts for the cellulase enzymes include two catalytic carboxyl residues. One carboxyl residue acts as an acid to protonate the scissille glycosidic bond, and the other acts as a base. The hydrophobic face of each glucose unit interacts with an aromatic side chain on the active site cleft. Whereas, the hydroxyl groups of each glucose interacts with hydrophilic residues. Most glycosyl hydrolase enzymes, that depolymerize polysaccharide molecules, share these structural features in common.

Highly thermostable cellulase enzymes are secreted by the cellulolytic thermophile *Acidothermus cellulolyticus*. These enzymes are disclosed in U.S. Pat. Nos. 5,110,735, 5,275,944, and 5,536,655 which are incorporated by reference as though fully set forth herein. This bacterium was isolated, in an acidic thermal pool at Yellowstone National Park, from decaying wood, and it is on deposit with the American Type Culture Collection under ATCC accession no. 43068. The cellulase complex produced by this organism contains several different cellulase enzymes. These enzymes are resistant to end-product-inhibition from

cellobiose and are active over a broad pH range, including the pH range at which yeasts are capable of fermenting glucose to ethanol. A novel endoglucanase, known as E1, is secreted by *Acidothermus cellulolyticus* into the growth medium. This enzyme is disclosed in U.S. Pat. No. 5,275,944. E1 endoglucanase exhibits a specific activity of 40 μmole glucose released from carboxymethylcellulose/min/mg protein.

In the prior art, it has been suggested to augment or replace costly naturally-occurring fungal cellulases with recombinant enzymes, which are useful in the digestion of cellulose. United States Pat. No. 5,536,655, incorporated herein, has disclosed that E1 endoglucanase is a candidate for recombination because the gene encoding E1 has been characterized, cloned and expressed in heterologous microorganisms. A new modified E1 endoglucanase enzyme has also been purified and four peptide sequences have been isolated. These four sequences include the signal, catalytic domain ("cd"), linker, and cellulose binding domains ("CBD") of the peptide. In U.S. Pat. No. 5,536,655 SEQ ID NO: 3 a single 521 amino acid linear-strand peptide is disclosed and contains, inter alia, the E1cd portion of the enzyme. Variants in the E1cd may be generated, through site-directed-mutagenesis of the E1 nucleotide sequence for translation, into a protein having an increase in catalytic activity over the wild-type E1. Information gained from the x-ray crystallographic structure of E1, Sakon, J., *et al.*, Crystal Structure of Thermostable Family 5 Endocellulase E1 from Acidothermus cellulolyticus in Complex with Cellotetraose, *Biochemistry*, Vol. 35, No. 33, 10648-10660, 1996, is useful in the selection of several amino acid sites, for replacement with non-native amino acids of varying chemistry. However, no replacements resulting in an increase in an increase in catalytic activity have been identified. Enhancement in the catalytic activity of E1, or glycosyl hydrolases in general, are needed to improve the cost efficiency of a process for the conversion of pretreated biomass to ethanol. Thus, in view of the foregoing considerations, there is an apparent need for variant endoglucanases having enhanced catalytic activity on cellulose derived substrates.

Disclosure of the Invention.

It is therefore an object of the invention to provide variant cellulase enzymes characterized by an improvement, over the wild-type enzyme, in the catalytic digestion of cellulose substrates.

Another object of the invention is to increase the specific activity of the E1 endoglucanase on the cellulose in pretreated biomass substrates.

It is yet another object of the invention to provide a method for increasing the specific activity on an insoluble substrate of a glycosyl hydrolase which is a structural analogue to E1 endoglucanase by replacing an active site glycosyl-stabilizing amino acid residue with a residue which does not strongly bind the disaccharide product from leaving the active site, i.e., which does not strongly bind the disaccharide product in the active site.

The foregoing specific objects and advantages of the invention are illustrative of those which can be achieved by the present invention and are not intended to be exhaustive or limiting of the possible advantages which can be realized. Thus, those and other objects and advantages of the invention will be apparent from the description herein or can be learned from practicing the invention, both as embodied herein or as modified in view of any variations which may be apparent to those skilled in the art.

Briefly, the invention provides a method for making a glycosyl hydrolase characterized by an increase in catalytic activity on an insoluble substrate, comprising replacing an active site associated glycosyl-stabilizing amino acid of the hydrolase with an amino acid, the replacing amino acid not strongly binding a disaccharide product in the active site, yet not adversely effecting enzymatic activity, and a method for making a glycosyl hydrolase characterized by an increase in catalytic activity on a soluble substrate, comprising replacing a hydrophobic substrate binding amino acid of the hydrolase with a positively charged amino acid. The invention further provides a glycosyl hydrolase, comprising Y42R (SEQ. ID NO:1), W82R (SEQ. ID NO:2), or Y245G (1) (SEQ. ID NO:3) and the DNA sequences encoding the enzymes.

Additional advantages of the present invention will be set forth in part in the description that follows, and will be obvious from that description or can be learned from practice of the invention.

Brief Description of the Drawings.

The accompanying drawing, which is incorporated in and which constitutes a part of the specification, illustrates at least one embodiment of the invention, and together with the description, explains the principles of the invention.

Figure 1 is a graphic representation of the Connolly surface rendering of the E1 endoglucanase Y245G mutation showing, as represented by the circular white spaces, the location of the cellodextrin substrate. The figure-eight-shaped-white-space, adjacent the +2 location, represents the location where the glycine for tryptophan substitution has been made in accordance with one example of the invention.

Best Mode for Carrying Out the Invention.

Unless specifically defined otherwise, all technical or scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described.

The sequence listings herein are variants of U.S. Pat. No. 5,536,655 SEQ ID NO: 3 having mutation replacements made for use according to the invention. Any reference which refers or relates a sequence herein includes the conservatively modified variants thereof.

"Structural analogs" means the structural analogs of E1 also benefiting from the E1 Y245G SEQ ID NO: 3 class of mutation, and include glycosyl hydrolases that provide stabilization for the leaving group, such as van der walls interaction, with an aromatic, sulfhydral, or hydrophobic side chain containing amino acid residues, and/or via hydrogen bonding interaction with amino acid side chains capable of hydrogen bonding to the sugar hydroxyl oxygen of hydrogen atoms. These analogous enzymes include both retaining and inverting enzymes.

Three examples for probing the possibility that the specific activity of an E1 glycosyl hydrolase can be increased, in a cellulose substrate, by site-directed mutagenesis ("SDM"), are provided. The first method describes replacing two hydrophobic surface-binding amino acid residue of the enzyme, such as residues tryptophan 42 and tyrosine 82 disclosed in U.S. Pat. No. 5,536,655 SEQ ID NO: 3 with a positively charged residue, such as is arginine (referenced herein as SEQ ID NO:1 W42R; and SEQ. ID NO:2 Y82R, respectively).

The second method includes replacing an active-site glycosyl-stabilizing amino acid residue of the enzyme, such as residue tyrosine 245 disclosed in U.S. Pat. No. 5,536,655 SEQ ID NO: 3 with a residue, such as glycine (referenced herein as SEQ. ID NO:3 Y245G), alanine,

valine, or serine, not strongly retarding cellobiose from leaving the active-site. Glycosyl hydrolase structural analogs of E1Y245G are set forth in Table 1. For example, in the Table, for the PDB code enzyme 1A3H (Brookhaven Data Base, Brookhaven National Laboratories) a replacement of Trp39 with Gly would remove Van der Waals stabilization of cellobiose and, it is believed that, it may also cause Gln 180 to adopt the non-native configuration in which it would be unable to hydrogen bond with cellobiose, the result being that cellobiose (the reaction product) would not strongly bind in the active-site, in the same manner as in the replacement made according to the E1Y245G example.

Table 1.

PDB code of Glycosyl Hydrolase Enzymes Structurally Related to E1	Mutation Sites: E1 Tyr245 Analog	Mutation Sites: E1 Gln247 Analog
1A3H	Trp39	Gln180
1BQC	Trp171	Gln169
1CEN	Trp212	Gln16, Asp319
1CZ1	Phe229, Phe258	
1EDG	Trp259, Trp181	
1EGZ		Gln172, Gln173, Lys205
2MAN	Trp30	

EXAMPLES

Various mutagenesis kits for SDM are available to those skilled in the art and the methods for SDM are well known. Three to four mutations were made for each E1 site W42, Y82, and Y245, including Ala, Gly, Glu, and Arg. The examples below illustrate a process for making and using these enzymes.

The *QuickChange* SDM kit, a trademark of Stratagene, San Diego, CA., was used to make point mutations, switch amino acids, and delete or insert amino acids in SEQ ID NO: 3 of U.S. Pat. No. 5,536,655. The *QuickChange* SDM technique was performed using a thermo-tolerant *Pfu* DNA polymerase, which replicates both plasmid strands with high fidelity, and without displacing the mutant oligonucleotide primers. The procedure used a polymerase chain reaction ("PCR") to alter the cloned E1 DNA (SEQ. ID NO: 6 of U.S. Pat. No. 5,536.655). The basic procedure used a super-coiled, double-stranded DNA (dsDNA) vector, with an insert of

interest, and two synthetic oligonucleotide primers containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, extend during temperature cycling by means of a *Pfu* DNA polymerase. On incorporation of the oligonucleotide primers, a mutated plasmid containing staggered nicks was generated. Following temperature cycling, the product was treated with the restriction enzyme, *Dpn*I. The *Dpn*I endonuclease (target sequence: 5'-(6-methyl)GATC-3') was specific for methylated and hemimethylated DNA and was used to digest the parental DNA template, and to select for mutation-containing newly synthesized DNA. The nicked vector DNA, incorporating the desired mutations, was then transformed into *E. coli*. The small amount of starting DNA template, required to perform this method, the high fidelity of the *Pfu* DNA polymerase and the low cycle number all contributed to a high mutation efficiency, and a decrease in the potential for random mutations during the reaction.

Template DNA (pBA100) was constructed using a 2.2 kb *Bam* H1 fragment, carrying most of the El gene including its native promoter, which functions in either *E. coli* or *S. lividans*, and approximately 800 kb of upstream sequence was sub-cloned into pUC19. The downstream *Bam* HI site cleaved the El coding sequence, at a point such that the protein was genetically truncated near the beginning of the linker peptide. Thus, the construct encoded a protein which included a signal peptide, the N-terminal cd and the first few amino acids of the C-terminal linker.

Using knowledge of the amino acid sequence of the crystalline Elcd structure, which was produced by papain cleavage of the holo-El protein, two different tandem translation terminator codons were introduced into the coding sequence, in frame, with the last amino acids present in the E1cd crystal structure. The 2.2 kb *Bam* HI fragment, named pBA100, in pUC19, containing the tandem stop codons, served as a template for the following mutagenesis reactions.

The three target sites of U.S. Pat. No. 5,536,655 SEQ ID NO: 3 selected for mutagenesis were W42, Y82, and Y245. Four or five pairs of mutagenic oligonucleotides were designed for each target site, such that 4 or 5 different amino acid substitutions would be created at each of the target sites. Both strands, of the template molecule, were copied and mutagenized during the invitro DNA synthesis reaction using the *QuickChange* In Vitro Mutagenesis kit (Strata Gene, San Diego, CA). The two mutagenic oligonucleotides were completely complementary to each

other, but they differed by one or more nucleotide from the template DNA strands. Each mutagenic oligonucleotide was designed, such that the nucleotides to be changed were located near the center of the oligonucleotide sequence, with approximately equal lengths of complementary sequence stretching out in both the 5′ and 3′ directions from the site of mutagenesis. Typically, mutagenic oligonucleotides were 26-30 nucleotides in length, but were sometimes longer due to considerations surrounding the melting temperature ("T_m"). The T_m was critical in the design of the mutagenic oligonucleotides because the oligonucleotides used in mutagenesis reactions required a T_m at least 10°C higher than the temperature for the DNA synthesis reaction (68°C). Accordingly, the effective mutagenic oligonucleotides required a T_m of at least 78°C.

Template DNA from *E. coli* XL1-blue cells, transformed with *Dpn*l treated mutagenized-DNA, was prepared for sequencing using the QIAprep-spin plasmid purification mini-prep procedure, provided by Qiagen, Inc. The transformed XLl-blue cells were grown over-night in 5 mL of LB broth with 100 μg/mL ampicillin. Cells were separated, by centrifugation, and the plasmid was isolated. Presence of the 2.2 kb insert was confirmed by digestion with *Bam*H1, followed by agarose electrophoresis. Transformants, having insert containing DNA, were precipitated in ethanol and then PEG. The DNA template concentration was adjusted to 0.25 μg/μL and the DNA was sequenced using procedures, which are well known in the art.

Transformed *E. coli* XL1/blue cells were cultured over-night at 37°C on LB plates containing 100 μg/mL ampicillin. A single colony was then used to inoculate 200 mL of LB broth, containing 100 μg/mL ampicillin in a 500 mL baffled Erlenmeyer flask. This organism was grown in a reciprocating incubator at 250 rpm, for 16-20 hours, at 37°C. This culture was used to inoculate a 10L BioFlow 3000 Chemostat, New Brunswick Scientific, New Brunswick New Jersey. The culture medium comprised LB broth, 100 μg/mL ampicillin, and 2.5% filter sterilized glucose. The pH, temperature, agitation rate, and dissolved oxygen parameters were maintained throughout the fermentation. The pH was controlled at 6.8 using a 2M potassium hydroxide solution. Temperature was controlled at 30°C, in order to prevent the formation of inclusion bodies. The agitation rate was 250 RPM. The dissolved oxygen polarographic probe was calibrated using nitrogen (0% activity at 4.0 L/min.) and house air (100% activity at 4.0

L/min). An oxygen and air mixture was used to maintain the dissolved oxygen tension at 20%. The cells were cultured 24-28 hours, which typically resulted in a maximum optical density of between 15-20. The cells were then harvested in a continuous centrifuge at 25,000 rpm.

Fifty grams of cells (wet/wt.) were added to the chamber of a stainless steel bead beater containing 200g of 0.1 mm glass beads, and 200mL of 20mM Tris, pH 8.0, buffer. Cell lysis was carried out for 5 min in the bead beater, while the chamber was chilled with ice. The contents of chamber was diluted two-fold, with buffer, and divided into centrifuge bottles (250 mL). The cell debris was removed by centrifugation at 13.000 rpm, 4°C, for 25 min. The supernatant was decanted, the pellet suspended in buffer, and the cells were milled and separated by centrifugation.

Two procedures were used in the initial purification of the enzyme(s). In the first, the supernatants were pooled and brought to 0.5M $(NH_4)_2SO_4$. The supernatant was divided, into 250 mL centrifuge bottles, and heated in a 65°C water bath, for 50 min, in order to denature non-E1 (i.e., *E. coli)* protein. Precipitated proteins were separated at 4°C by centrifugation at 13,000 rpm, for 25 min.. The supernatant was then filtered, through a glass fiber filter pad, prior to the chromatography step. An improved purification procedure resulted in a substantial reduction in the overall processing-time, but retained an equivalent yield of protein. This procedure involved lysing the cells, using the mill, combining the supernatants, and diluting the combined supernatant with 20 mM Tris, pH 8.0, buffer until the conductivity of the supernatant was less than 2000 µS/cm. The resulting material was separated, with an expanded-bed-adsorption-chromatography system using DEAE packing, in a Pharmacia StreamLine column.

Two methods were developed for the subsequent purification of the mutant E1 enzymes from the *E. coli* XL1/blue cell lysates described above. The original protocol involved a substantial amount of sample preparation prior to purification. An improved procedure was subsequently developed using new chromatography resins, which eliminated the need for much of the sample preparation and clarification of the cell lysate.

The original purification protocol comprised the following steps. The cell lysate, which contained 0.5 M $(NH_4)_2SO_4$, was loaded on a Pharmacia preparative chromatography column which had been packed with a 500 mL bed volume of Pharmacia Fast Flow, low substitution Phenyl Sepharose media. A Pharmacia BioPilot system was used to control chromatography.

After the cell lysate was loaded, the column was washed with three to five volumes of 20mM Tris, pH 8.0, buffer containing 0.5 M $(NH_4)_2SO_4$, at a flow rate of 0.50 DL/min, after which the recombinant E1 enzyme(s) ("rEl") was eluted with 3.2 column volumes, descending linear gradient, to zero-percent salt of 20 mM Tris, pH 8.0, buffer. The rE1 eluted in fractions resulting from approximately zero percent salt. These fractions were combined, and dialyzed against 20 mM Tris, pH 8.0, buffer for 12 hours. The dialyzed-concentrated-protein was subjected to anion-exchange-chromatography in a Pharmacia Q-Sepharose HiLoad 16/10 high performance column. The enzyme was loaded in 20 mM Tris, pH 8.0, buffer, and was eluted by a shallow linear gradient (22 column volumes) using the same buffer with 0.5 M NaCl. Most of the rEl mutant enzyme(s) eluted at 150mM NaCl. The active fractions were then combined, concentrated, and loaded in a Pharmacia Superdex 200 HiLoad prep grade column, at a 0.5 mL/min. flow rate in 20 mM acetate, pH 5.0, buffer with 100mM NaCl. The rE1 enzymes eluted as a single-symmetrical-peak, which is indicative of a highly homogenous compound. The purity of the rE1 enzyme(s) was confirmed with SDS-PAGE using Novex pre-cast 8-15% gradient gels, and contained a single 40 kDa band. The protein concentrations were then determined based on absorbance at 280 nm using a molar extinction coefficient which had been calculated for each altered enzyme based on the individual replacement amino acid.

The improved method eliminated the need for clarification of the supernatant after lysing the cells. The cell lysate, which had been adjusted to a conductivity of less than 2000 µS/cm, was loaded directly onto a Pharmacia StreamLine column packed with StreamLine DEAE (a weak anion-exchanger) fluidized at a flow rate of 15 mL/min with 20 mM Tris, pH 8.0, buffer. After the column matrix was washed free of the cell debris, and the UV absorbance returned close to zero, the flow was reversed to a down-flow orientation, and the proteins were eluted using a linear gradient of 20mM Tris, 1M NaCl, pH 8.0, buffer. Active fractions were pooled, and ammonium sulfate was added to a final concentration of 0.5M. These samples were then loaded on a Phenyl Sepharose HiLoad column. After the column was washed, with 3-5 column volumes of the starting buffer, the rE1 enzyme(s) was eluted, by a 3.2 column-volume descending linear gradient, to zero percent salt in 20 mM Tris, pH 8.0, buffer. The final purification step, and buffer exchange, was made using a Superdex 200, HiLoad prep-grade-column with a flow rate of 0.5 mL/min., in 20 mM acetate, pH 5.0, buffer with 100mM NaCl.

Mutant rE1 enzymes eluted as single symmetrical peaks, indicating a high level of homogeneity. The protein concentrations were then determined as described above.

Solid-phase immunology methods were used to detect the expressed enzyme. Immunoblots and Western blots were used to verify the presence of E1 and E1 mutant enzymes. For immunoblots, 2 μL of a chromatography sample fraction was applied to nitrocellulose and allowed to air dry. For Western blots, 3-5 μg of protein was added to each lane, and the proteins were subjected to electrophoresis. A monoclonal antibody specific for E1 was then added after the proteins had been blotted to the nitrocellulose. This was followed by the addition of a goat anti-mouse-IgG alkaline phosphate-labeled antibody. Bound E1 was visualized by the precipitation of the substrate.

The Michaelis constant ("K_m") and maximal rate ("V_{max}") for each enzyme preparation were determined from the rates of cellobiose production, at different cellotriose concentrations. Replicate assay mixtures containing 5mM acetate buffer, pH 5.0, 10μg/mL BSA, and cellotriose ranging from 0.0793mM (0.04 mg/mL) to 1.9825 mM (1.0 mg/mL) were prepared. Each assay mixture was pre-incubated at 50°C for 10 min, prior to the addition of 0.00272 μM (0. 1092 μg/mL) enzyme, which was also made up in 5mM acetate buffer with 10 μg/mL BSA. The final assay volume was 1.0mL.

At set-time intervals, an aliquot of the reaction mixture was pulled and immediately analyzed for the release of cellobiose using a Dionex DX300 chromatography system, and a Dionex PAD2 pulsed amperometric detector having a gold working electrode. The response of this detector was optimized for the detection of carbohydrates, using a waveform defined by the following time and potential settings: t_1 = 420 msec; E1 = +0.05 V; t_2= 180 msec; E_2= +0.75 V; t_3 = 360 msec; and E_3 = -0.15 V. Separation of the reaction products, from the substrate, was achieved on a Dionex CarboPac PA-1 analytical (4 x 250 mm) column equipped with CarboPac PA-1 (4 x 50 mm) guard column, 500 mM sodium hydroxide eluent, and a flow rate of 1.5 mL/min. The amount of cellobiose, present for each time-point-sample, was quantified by comparing the area of the cellobiose peak against a linear calibration curve. The kinetic constants were determined with a double-reciprocal-plot, where the reciprocal of the rate of cellobiose produced was plotted as a function of the inverse of the substrate concentration. This resulted in a straight line function having an intercept of $1/V_{max}$ and a slope of K_m/V_{max}.

All diafiltration saccharification assays ("DSA") were carried out at 50°C in 20mM, pH 5.0, sodium acetate buffer containing 0.02% sodium azide. Substrate loading, for each assay, comprised 104mg (dry wt.) of pretreated-yellow-poplar ("PYP"). This weight was equal to a load having 4.7% biomass and a 3.2% cellulose. The substrate was ground to a maximum particle size of between 10 and 500 microns. Selected enzymes, such as the wild-type or mutant *A. cellulolyticus* E1 catalytic domain, were loaded at 56.4 nanomoles enzyme/g cellulose. Each assay mixture further included 487 nanomoles of *T reesei* cellobiohydrolase (CBH 1) enzyme/g cellulose, which resulted in an enzymatic solution of 10% endoglucanase and 90% cellobiohydrolase. The endoglucanase proportion in the mixture was high enough to provide a readily-measurable activity, but was sufficiently below an optimal endoglucanase concentration, which causes sugar release and synergism to make the results highly sensitive to differences in endoglucanase activity.

The temperature optima for maximum activity was determined for each E1 mutant using *p*-nitrophenol-β-D-cellobioside as the substrate in a 20mM acetate, 100mM NaCl, pH 5.0, buffer. Equivalent concentrations of enzyme were used (0.4 μg/mL) in a 30 min assay at various temperatures. After a 30 min incubation period, the reactions were stopped with the addition of 2mL 1M Na_2CO_3 and the amount of *p*-nitrophenolate anion released was measured by absorbance at 410 nm. The temperature optima for the mutants claimed was found to be essentially identical to that of the native E1.

While the PCR technique is well known in the art and commonly performed with reagents packaged in kit form, the following modifications provided nucleotide substitutions at all targeted sites, which are identified in the Table 2 below. Good annealing of the DNA template and primers was critical. The T_m for this process was a function of the length of the oligonucleotide, the concentration of monovalent cations, and the GC content of the oligonucleotide. The T_m was calculated according to the formula: $T_m = 81.5 + 16.6(\log[Na+]) + 0.41(\%\ G+C) - (675 / N) - \%$ mismatch, where N is the primer length in base pairs, and [Na+] is the sodium ion concentration. The T_m increased with an increase in the GC content, salt concentration, and oligonucleotide length. Because the E1 sequence is very GC-rich (62.8%), relatively short mutagenic oligonucleotides were used (i.e., 26-30 bases). However, in some situations because of the relatively AT-rich segment of DNA around a site (i.e., lower T_m), such

as was the case for the Y82 mutations, longer mutagenic oligonucleotides (38 bases) were synthesized in order to obtain an oligonucleotide having a suitably high T_m. Table 2 illustrates the mutations in SEQ ID NO:6 US PAT. NO 5,536,655 which translated into the rE1 enzymes demonstrating an increase in activity over the native protein of SEQ ID NO:3 US PAT. NO 5,536,655. Changing the codon(s) to reflect an alanine, valine, or serine replacement can be made in the same of similar manner, and the codons for these amino acids are well known.

Table 2.

E1 Mutation Target Site SEQ ID NO:3 US PAT. NO 5,536,655	Insert DNA Sequence From PCR Mutation SEQ ID NO:6 US PAT. NO 5,536,655
E1W42 NATIVE	GTGCACGGTC TCTGGTCACG CGACTACCG
E1W42R	GTGCACGGTC TCCGGTCACG CGACTACCG
E1Y82 NATIVE	GC CGAACAGCAT CAATTTTTAC CAGATGAATC AGGACC
E1Y82R	GC CGAACAGCAT CAATTTTCGC CAGATGAATC AGGACC
E1Y245 NATIVE	CGCGACGAGC GTCTACCCGC AGACGTGG
E1Y245G	CGCGACGAGC GTCGGCCCGC AGACGTGG

Industrial Applicability.

The mutant E1 enzymes and one native E1cd were purified using the purification methods described above. Purification of the mutant enzymes destined for kinetic analysis was necessary because any precise comparison of specific activity required knowledge of the enzyme(s) concentration. For this reason, a determination of the molar extinction coefficients of the recombinant enzymes was made by considering the specific change in the amino acid compositions. Although all active mutant E1 enzymes behaved similarly during purification, some mutant enzymes showed a substantial departure from the E1cd behavior on anion exchange chromatography. All transformed strains of *E. coli* examined were found to produce adequate levels of mutant E1 enzymes (i.e., approximately 0.5 to 1 mg/10 L culture).

Ten-Liter cultures of the transformed *E. coli*. expressing active enzymes. were grown. and each mutant enzyme was purified to homogeneity using an improved three-step column

chromatographic method. The purified rE1 endoglucanase enzymes (including the E1cd control) were characterized for activity on cellotriose and PYP.

Michaelis-Menten kinetics of the mutant El enzymes and the native enzyme were determined. As a result, it was concluded that the W42R (SEQ ID NO:1) and Y82R (SEQ. ID NO:2) amino acid substitutions at sites W42 and Y82 of U.S. Pat. No. 5,536,655 SEQ ID NO: 3 improved the catalytic activity for this soluble substrate.

Cellotriose kinetics for the E1 mutations are show in the Table 3 below. In the case of cellotriose hydrolysis, mutations which increased K_m (indicating probable decreases in strength of substrate binding), also displayed an increases in velocity. Thus, the arginine substitutions at sites W42 and Y82 resulted in the highest V_{max} values observed, about 15% and 75% higher than that of the native enzyme, respectively.

Table 3.

Enzyme/Mutant	Km(mM)	Vmax(uM/min.)
E1 NATIVE	0.35	0.86
W42R	0.61	0.99
Y82R	0.69	1.5
Y245G	0.48	0.85

These mutant E1 enzymes were also tested for activity on pretreated yellow poplar using the diafiltration saccharification assay. Baker, J.O., *et al.*, Use of a New Membrane-Reactor Saccharification Assay to Evaluate the Performance of Cellulases Under Simulated SSF Conditions, *Applied Biochemistry and Bioengineering*, 1997, Vol. 63-65, 585-595. This assay tested the ability of the modified El enzymes to hydrolyze an insoluble substrate in combination with *T. reesei* cellobiohydrolase (CBH 1). This test has the advantage of taking cellulose hydrolysis to the 90% level, under conditions consistent with simultaneous saccharification fermentation, which is a desirable use for the cellulase enzymes according to the examples herein.

Ten-L cultures of the transformed *E coli* expressing active enzymes were grown and each mutant enzyme was purified to homogeneity using an improved three-step column

chromatographic method. The purified E1 endoglucanase enzymes (including the E1 control) underwent DSA on cellulose. In Table 4, the results for the E1 mutations, having at least native activity, are shown.

Table 4.

ENZYME/MUTANT	% SACCHARIFICATION OF PYP / 96HOURS
E1 NATIVE	44.5 (+/- 0.5%)
W42R	46
Y82R	45.3
Y245G	50.5

Although 3 to 4 mutation were made for each E1 site W42, Y82, and Y245, including Ala, Gly, Glu, Gln, and Arg, only three variants demonstrated no loss in native activity on insoluble substrates relative to the native enzyme. These E1 variants were identified as W42R, Y82R, and Y245G. Only the E1Y245G (U.S. Pat. No. 5,536,655 SEQ. ID NO:3) variant showed a significantly greater catalytic activity over native E1. DSA testing revealed that the glycine mutant enzyme (Y245G) demonstrated a 12% (+/- 1.0%) improvement in DSA catalytic activity. This increase is explained by a decrease in cellobiose binding, and thus cellobiose end-product-inhibition at site Y245. To confirm this result, a second preparation of E1Y245G was produced from the transformed *E. coli* stock. This mutant E1 also showed substantial increase in DSA activity over the native enzyme. i.e., 9.5% (+/- 1.0%).

Results suggesting that the relief of inhibition by cellobiose is a factor in enhanced biomass hydrolysis, with the E1Y245G mutant, are supported from the following observations: (1) addition to the DSA enzyme cocktail of sufficient β-D-glucosidase, to reduce the cellobiose concentration the assay reactor below the level of HPLC selectability, has the effect of abolishing most of the difference in performance between native and mutant E1; and (2) K_i values for inhibition of hydrolysis of 4-β-D-cellobioside (MUC) by native and mutant E1 indicate that the mutant catalytic domain binds cellobiose 15 times less tightly than does the native enzyme, i.e., an increase in K_i from 2 to 30 mM cellobiose. The decrease in apparent binding energy is 1.7 kcal/mol.

The foregoing description is considered as illustrative only of the principles of the

invention. Furthermore, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and process shown as described above. Accordingly, all suitable modifications and equivalents may be resorted to falling within the scope of the invention as defined by the claims which follow.

Claims

We claim:

1. A method for making a glycosyl hydrolase characterized by an increase in catalytic activity on an insoluble substrate, comprising replacing an active site associated glycosyl-stabilizing amino acid of the hydrolase with an amino acid, the replacing amino acid not strongly binding a disaccharide product in the active site.

2. The method of claim 1 wherein the glycosyl-stabilizing amino acid comprises tyrosine 245 of U.S. Pat. No. 5,536,655 SEQ. ID NO: 3 and the replacing amino acid comprises glycine.

3. The method of claim 1 wherein the glycosyl hydrolase is selected from a group consisting of a PDB code identification 1AEH, 1BQC, 1CEN, 1CZ1, 1EDG, 1EGZ, and 2MAN, the glycosyl-stabilizing amino acid is selected from the group consisting of Trp39, Trp171, Trp212, Phe229 and Phe258, Trp259 and Trp181, and Trp30, respectively, and the replacing amino acid comprises glycine.

4. The method of claim 2 wherein the increase in catalytic activity of an insoluble substrate consisting essentially of cellulose in a pretreated-yellow-poplar hydrolysate comprises a 10% to 12% improvement in a diafiltration-saccharification-assay results over a native E1.

5. An E1 endoglucanase Y245G SEQ. ID NO: 3.

6. A glycosyl hydrolase comprising a structural analog to the E1 endoglucanase of claim 5.

7. A glycosyl hydrolase of claim 6 wherein the structural analog comprises an enzyme selected from a group consisting the a PDB code identification 1AEH, 1BQC, 1CEN, 1CZ1, 1EDG, 1EGZ, and 2MAN, and a glycosyl-stabilizing amino acid selected from the group consisting of Trp39, Trp171, Trp212, Phe229 and Phe258, Trp259 and Trp181, and Trp30, respectively, and the replacing amino acid comprises glycine.

8. A DNA sequence encoding the endoglucanase of claim 5.

9. A DNA sequence encoding the endoglucanase of claim 6.

10. A DNA sequence encoding the endoglucanase of claim 7.

11. The use of the endoglucanase of claim 5 as a catalyst for cellulose hydrolysis.

12. The use of the endoglucanase of claim 6 as a catalyst for cellulose hydrolysis.

13. The use of the endoglucanase of claim 7 as a catalyst for cellulose hydrolysis.

14. A method for making a glycosyl hydrolase characterized by an increase in catalytic

activity on a soluble substrate, comprising replacing a hydrophobic substrate binding amino acid of the hydrolase with a positively charged amino acid.

15. The method of claim 14 wherein the glycosyl hydrolase is E1 and the increase in catalytic activity of a substrate consisting essentially of cellotriose comprises a V_{max} in a range of 15% to 75% greater than native E1.

16. An E1 endoglucanase comprising Y42R SEQ. ID NO:1.

17. An E1 endoglucanase comprising W82R SEQ. ID NO:2.

18. A DNA sequence encoding the endoglucanase of claim 16.

19. A DNA sequence encoding the endoglucanase of claim 17.

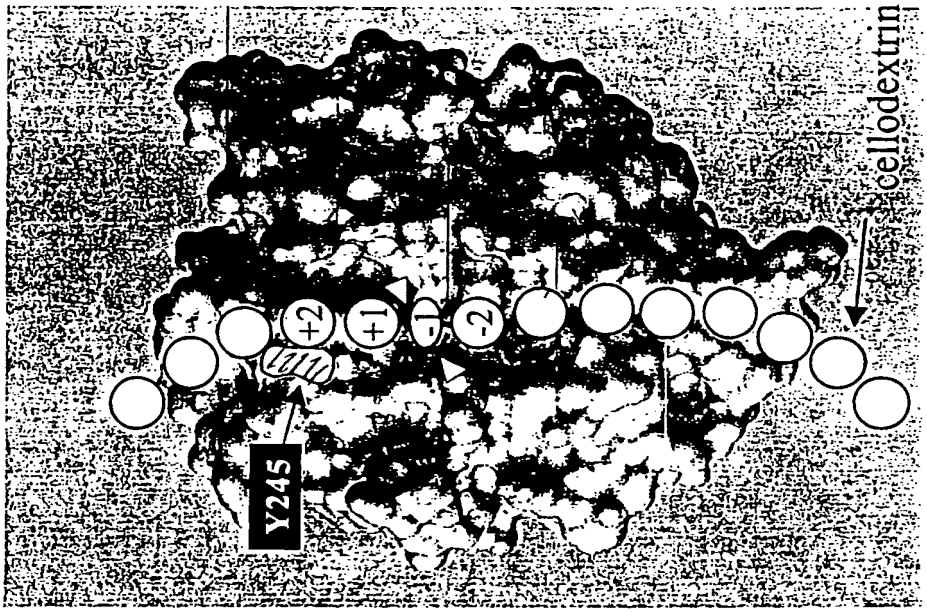

FIGURE 1

SEQUENCE LISTING

<110> National Renewable Energy Laboratory

<120> Endoglucanase Variants Y245G, Y82R, and W42R

<130> NREL PCT 99-38

<140>
<141>

<150> USPROV 60/134,917
<151> 1999-05-19

<160> 3

<170> PatentIn Ver. 2.1

<210> 1
<211> 521
<212> PRT
<213> Acidothermus Cellulolyticus

<220>
<221> MUTAGEN
<222> (42)

<220>
<221> DOMAIN
<222> (42)..(404)

<300>
<301> Thomas et al, Steven R.
<302> Gene Coding For The E1 Endoglucanase, SEQ ID: 3
<310> 5,536,655
<311> 1994-07-15
<312> 1996-07-16
<313> 42 TO 245

<400> 1

```
Ala Gly Gly Gly Tyr Trp His Thr Ser Gly Arg Glu Ile Leu Asp Ala
  1               5                  10                  15

Asn Asn Val Pro Val Arg Ile Ala Gly Ile Asn Trp Phe Gly Phe Glu
             20                  25                  30

Thr Cys Asn Tyr Val Val His Gly Leu Arg Ser Arg Asp Tyr Arg Ser
         35                  40                  45
```

```
Met Leu Asp Gln Ile Lys Ser Leu Gly Tyr Asn Thr Ile Arg Leu Pro
     50                  55                  60

Tyr Ser Asp Asp Ile Leu Lys Pro Gly Thr Met Pro Asn Ser Ile Asn
 65                  70                  75                  80

Phe Tyr Gln Met Asn Gln Asp Leu Gln Gly Leu Thr Ser Leu Gln Val
                 85                  90                  95

Met Asp Lys Ile Val Ala Tyr Ala Gly Gln Ile Gly Leu Arg Ile Ile
            100                 105                 110

Leu Asp Arg His Arg Pro Asp Cys Ser Gly Gln Ser Ala Leu Trp Tyr
        115                 120                 125

Thr Ser Ser Val Ser Glu Ala Thr Trp Ile Ser Asp Leu Gln Ala Leu
    130                 135                 140

Ala Gln Arg Tyr Lys Gly Asn Pro Thr Val Val Gly Phe Asp Leu His
145                 150                 155                 160

Asn Glu Pro His Asp Pro Ala Cys Trp Gly Cys Gly Asp Pro Ser Ile
                165                 170                 175

Asp Trp Arg Leu Ala Ala Glu Arg Ala Gly Asn Ala Val Leu Ser Val
            180                 185                 190

Asn Pro Asn Leu Leu Ile Phe Val Glu Gly Val Gln Ser Tyr Asn Gly
        195                 200                 205

Asp Ser Tyr Trp Trp Gly Gly Asn Leu Gln Gly Ala Gly Gln Tyr Pro
    210                 215                 220

Val Val Leu Asn Val Pro Asn Arg Leu Val Tyr Ser Ala His Asp Tyr
225                 230                 235                 240

Ala Thr Ser Val Tyr Pro Gln Thr Trp Phe Ser Asp Pro Thr Phe Pro
                245                 250                 255

Asn Asn Met Pro Gly Ile Trp Asn Lys Asn Trp Gly Tyr Leu Phe Asn
            260                 265                 270

Gln Asn Ile Ala Pro Val Trp Leu Gly Glu Phe Gly Thr Thr Leu Gln
        275                 280                 285

Ser Thr Thr Asp Gln Thr Trp Leu Lys Thr Leu Val Gln Tyr Leu Arg
    290                 295                 300
```

```
Pro Thr Ala Gln Tyr Gly Ala Asp Ser Phe Gln Trp Thr Phe Trp Ser
305                 310                 315                 320

Trp Asn Pro Asp Ser Gly Asp Thr Gly Gly Ile Leu Lys Asp Asp Trp
                325                 330                 335

Gln Thr Val Asp Thr Val Lys Asp Gly Tyr Leu Ala Pro Ile Lys Ser
            340                 345                 350

Ser Ile Phe Asp Pro Val Gly Ala Ser Ala Ser Pro Ser Ser Gln Pro
        355                 360                 365

Ser Pro Ser Val Ser Pro Ser Pro Ser Pro Ser Pro Ser Ala Ser Arg
    370                 375                 380

Thr Pro Thr Pro Thr Pro Thr Pro Thr Ala Ser Pro Thr Pro Thr Leu
385                 390                 395                 400

Thr Pro Thr Ala Thr Pro Thr Pro Thr Ala Ser Pro Thr Pro Ser Pro
                405                 410                 415

Thr Ala Ala Ser Gly Ala Arg Cys Thr Ala Ser Tyr Gln Val Asn Ser
            420                 425                 430

Asp Trp Gly Asn Gly Phe Thr Val Thr Val Ala Val Thr Asn Ser Gly
        435                 440                 445

Ser Val Ala Thr Lys Thr Trp Thr Val Ser Trp Thr Phe Gly Gly Asn
    450                 455                 460

Gln Thr Ile Thr Asn Ser Trp Asn Ala Ala Val Thr Gln Asn Gly Gln
465                 470                 475                 480

Ser Val Thr Ala Arg Asn Met Ser Tyr Asn Asn Val Ile Gln Pro Gly
                485                 490                 495

Gln Asn Thr Thr Phe Gly Phe Gln Ala Ser Tyr Thr Gly Ser Asn Ala
            500                 505                 510

Ala Pro Thr Val Ala Cys Ala Ala Ser
        515                 520
```

<210> 2
<211> 521
<212> PRT
<213> Acidothermus cellulolyticus

<220>
<221> MUTAGEN
<222> (82)

<220>
<221> DOMAIN
<222> (42)..(404)

<400> 2

```
Ala Gly Gly Gly Tyr Trp His Thr Ser Gly Arg Glu Ile Leu Asp Ala
  1               5                  10                  15

Asn Asn Val Pro Val Arg Ile Ala Gly Ile Asn Trp Phe Gly Phe Glu
             20                  25                  30

Thr Cys Asn Tyr Val Val His Gly Leu Trp Ser Arg Asp Tyr Arg Ser
         35                  40                  45

Met Leu Asp Gln Ile Lys Ser Leu Gly Tyr Asn Thr Ile Arg Leu Pro
     50                  55                  60

Tyr Ser Asp Asp Ile Leu Lys Pro Gly Thr Met Pro Asn Ser Ile Asn
 65                  70                  75                  80

Phe Arg Gln Met Asn Gln Asp Leu Gln Gly Leu Thr Ser Leu Gln Val
                 85                  90                  95

Met Asp Lys Ile Val Ala Tyr Ala Gly Gln Ile Gly Leu Arg Ile Ile
            100                 105                 110

Leu Asp Arg His Arg Pro Asp Cys Ser Gly Gln Ser Ala Leu Trp Tyr
        115                 120                 125

Thr Ser Ser Val Ser Glu Ala Thr Trp Ile Ser Asp Leu Gln Ala Leu
    130                 135                 140

Ala Gln Arg Tyr Lys Gly Asn Pro Thr Val Val Gly Phe Asp Leu His
145                 150                 155                 160

Asn Glu Pro His Asp Pro Ala Cys Trp Gly Cys Gly Asp Pro Ser Ile
                165                 170                 175

Asp Trp Arg Leu Ala Ala Glu Arg Ala Gly Asn Ala Val Leu Ser Val
            180                 185                 190

Asn Pro Asn Leu Leu Ile Phe Val Glu Gly Val Gln Ser Tyr Asn Gly
        195                 200                 205
```

Asp Ser Tyr Trp Trp Gly Gly Asn Leu Gln Gly Ala Gly Gln Tyr Pro
210 215 220

Val Val Leu Asn Val Pro Asn Arg Leu Val Tyr Ser Ala His Asp Tyr
225 230 235 240

Ala Thr Ser Val Tyr Pro Gln Thr Trp Phe Ser Asp Pro Thr Phe Pro
245 250 255

Asn Asn Met Pro Gly Ile Trp Asn Lys Asn Trp Gly Tyr Leu Phe Asn
260 265 270

Gln Asn Ile Ala Pro Val Trp Leu Gly Glu Phe Gly Thr Thr Leu Gln
275 280 285

Ser Thr Thr Asp Gln Thr Trp Leu Lys Thr Leu Val Gln Tyr Leu Arg
290 295 300

Pro Thr Ala Gln Tyr Gly Ala Asp Ser Phe Gln Trp Thr Phe Trp Ser
305 310 315 320

Trp Asn Pro Asp Ser Gly Asp Thr Gly Gly Ile Leu Lys Asp Asp Trp
325 330 335

Gln Thr Val Asp Thr Val Lys Asp Gly Tyr Leu Ala Pro Ile Lys Ser
340 345 350

Ser Ile Phe Asp Pro Val Gly Ala Ser Ala Ser Pro Ser Ser Gln Pro
355 360 365

Ser Pro Ser Val Ser Pro Ser Pro Ser Pro Ser Pro Ser Ala Ser Arg
370 375 380

Thr Pro Thr Pro Thr Pro Thr Pro Thr Ala Ser Pro Thr Pro Thr Leu
385 390 395 400

Thr Pro Thr Ala Thr Pro Thr Pro Thr Ala Ser Pro Thr Pro Ser Pro
405 410 415

Thr Ala Ala Ser Gly Ala Arg Cys Thr Ala Ser Tyr Gln Val Asn Ser
420 425 430

Asp Trp Gly Asn Gly Phe Thr Val Thr Val Ala Val Thr Asn Ser Gly
435 440 445

Ser Val Ala Thr Lys Thr Trp Thr Val Ser Trp Thr Phe Gly Gly Asn
450 455 460

```
Gln Thr Ile Thr Asn Ser Trp Asn Ala Ala Val Thr Gln Asn Gly Gln
465                 470                 475                 480

Ser Val Thr Ala Arg Asn Met Ser Tyr Asn Asn Val Ile Gln Pro Gly
                485                 490                 495

Gln Asn Thr Thr Phe Gly Phe Gln Ala Ser Tyr Thr Gly Ser Asn Ala
            500                 505                 510

Ala Pro Thr Val Ala Cys Ala Ala Ser
        515                 520
```

<210> 3
<211> 521
<212> PRT
<213> Acidothermus cellulolyticus

<220>
<221> MUTAGEN
<222> (245)

<220>
<221> DOMAIN
<222> (42)..(404)

<400> 3

```
Ala Gly Gly Gly Tyr Trp His Thr Ser Gly Arg Glu Ile Leu Asp Ala
  1               5                  10                  15

Asn Asn Val Pro Val Arg Ile Ala Gly Ile Asn Trp Phe Gly Phe Glu
            20                  25                  30

Thr Cys Asn Tyr Val Val His Gly Leu Trp Ser Arg Asp Tyr Arg Ser
        35                  40                  45

Met Leu Asp Gln Ile Lys Ser Leu Gly Tyr Asn Thr Ile Arg Leu Pro
    50                  55                  60

Tyr Ser Asp Asp Ile Leu Lys Pro Gly Thr Met Pro Asn Ser Ile Asn
65                  70                  75                  80

Phe Tyr Gln Met Asn Gln Asp Leu Gln Gly Leu Thr Ser Leu Gln Val
                85                  90                  95

Met Asp Lys Ile Val Ala Tyr Ala Gly Gln Ile Gly Leu Arg Ile Ile
            100                 105                 110
```

```
Leu Asp Arg His Arg Pro Asp Cys Ser Gly Gln Ser Ala Leu Trp Tyr
        115                 120                 125

Thr Ser Ser Val Ser Glu Ala Thr Trp Ile Ser Asp Leu Gln Ala Leu
    130                 135                 140

Ala Gln Arg Tyr Lys Gly Asn Pro Thr Val Val Gly Phe Asp Leu His
145                 150                 155                 160

Asn Glu Pro His Asp Pro Ala Cys Trp Gly Cys Gly Asp Pro Ser Ile
                165                 170                 175

Asp Trp Arg Leu Ala Ala Glu Arg Ala Gly Asn Ala Val Leu Ser Val
            180                 185                 190

Asn Pro Asn Leu Leu Ile Phe Val Glu Gly Val Gln Ser Tyr Asn Gly
        195                 200                 205

Asp Ser Tyr Trp Trp Gly Gly Asn Leu Gln Gly Ala Gly Gln Tyr Pro
    210                 215                 220

Val Val Leu Asn Val Pro Asn Arg Leu Val Tyr Ser Ala His Asp Tyr
225                 230                 235                 240

Ala Thr Ser Val Gly Pro Gln Thr Trp Phe Ser Asp Pro Thr Phe Pro
                245                 250                 255

Asn Asn Met Pro Gly Ile Trp Asn Lys Asn Trp Gly Tyr Leu Phe Asn
            260                 265                 270

Gln Asn Ile Ala Pro Val Trp Leu Gly Glu Phe Gly Thr Thr Leu Gln
        275                 280                 285

Ser Thr Thr Asp Gln Thr Trp Leu Lys Thr Leu Val Gln Tyr Leu Arg
    290                 295                 300

Pro Thr Ala Gln Tyr Gly Ala Asp Ser Phe Gln Trp Thr Phe Trp Ser
305                 310                 315                 320

Trp Asn Pro Asp Ser Gly Asp Thr Gly Gly Ile Leu Lys Asp Asp Trp
                325                 330                 335

Gln Thr Val Asp Thr Val Lys Asp Gly Tyr Leu Ala Pro Ile Lys Ser
            340                 345                 350

Ser Ile Phe Asp Pro Val Gly Ala Ser Ala Ser Pro Ser Ser Gln Pro
        355                 360                 365
```

```
Ser Pro Ser Val Ser Pro Ser Pro Ser Pro Ser Pro Ser Ala Ser Arg
    370                 375                 380

Thr Pro Thr Pro Thr Pro Thr Pro Thr Ala Ser Pro Thr Pro Thr Leu
385                 390                 395                 400

Thr Pro Thr Ala Thr Pro Thr Pro Thr Ala Ser Pro Thr Pro Ser Pro
                405                 410                 415

Thr Ala Ala Ser Gly Ala Arg Cys Thr Ala Ser Tyr Gln Val Asn Ser
            420                 425                 430

Asp Trp Gly Asn Gly Phe Thr Val Thr Val Ala Val Thr Asn Ser Gly
        435                 440                 445

Ser Val Ala Thr Lys Thr Trp Thr Val Ser Trp Thr Phe Gly Gly Asn
    450                 455                 460

Gln Thr Ile Thr Asn Ser Trp Asn Ala Ala Val Thr Gln Asn Gly Gln
465                 470                 475                 480

Ser Val Thr Ala Arg Asn Met Ser Tyr Asn Asn Val Ile Gln Pro Gly
                485                 490                 495

Gln Asn Thr Thr Phe Gly Phe Gln Ala Ser Tyr Thr Gly Ser Asn Ala
            500                 505                 510

Ala Pro Thr Val Ala Cys Ala Ala Ser
        515                 520
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/13971

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :C12N 9/24, 9/26, 9/42
US CL :435/200, 201, 209
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/200, 201, 209

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,716,812 A (WITHERS et al) 10 February 1998, see entire document.	1-19
Y	US 4,760,025 A (ESTELL et al) 26 July 1988, see entire document.	1-19
Y	US 5,536,655 A (THOMAS et al) 16 July 1996, see entire document.	1-19

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
26 JULY 2000

Date of mailing of the international search report
17 AUG 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
L. BLAINE LANKFORD
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*